How is statistical significance connected with goodness-of-fit?

Dr Kateryna Zabarina,

Assistant professor, Chair of Statistics and Econometrics, FoES, UW

Dr Dorota Celińska-Kopczyńska,

Assistant professor, Institute of Informatics, MIM, UW

IV Kongres Statystyki Polskiej, 2-4 lipca 2024 r., Warszawa // IV Congress of Polish Statistics, July 2nd – 4th, Warsaw

UNIVERSITY OF WARSAW | Faculty of Economic Sciences

WYDZIAŁ MATEMATYKI, INFORMATYKI I MECHANIKI MIM

UNIVERSITY OF WARSAW

GUS

Agenda

1. Aim of the study
2. Theoretical issues
3. Data and covariates
4. Estimation results
5. What can single hypothesis tell us?
6. Simulation results
7. Implications and conclusions
8. Additional slides*

0. Introduction

- Previous research on agglomeration based mostly on aggregated data or didn't account any relationship between firms (e.g. Figueiredo et al., 2002; Dube et al., 2016; Credit, 2019)
- This research uses **modern** point pattern analysis technics – in particular – Gibbs processes, and models locations of firms accounting not only for space features, but also for an **unobservable interactions** – **competition for resources, clients** etc. (not knowledge spillovers as it is present in classical Marshallian or Jacobian agglomeration definition)
- Following Sweeney and Gómez-Antonio, **Hoover's definition** of agglomeration will be accepted (*"the existence of localisation economies"* (Sweeney & Gómez-Antonio, 2016, p. 258)), for which the point process modelling approach will be applied.
- Localisation economy will be treated as *Marshallian* type of agglomeration.
- Such consideration **does not** take into account any kind of spillovers, inputs or knowledge exchange (here, I try to follow Christ (2008), who names lack of knowledge flows as a feature of localisation/urbanisation economies) -> Sweeney & Gómez-Antonio 2021.

1. Aim of the study

- Research aims to check **how statistical significance of estimated parameters is connected with model's goodness-of-fit**.

1. Aim of the study

- Research aims to check **how statistical significance of estimated parameters is connected with model's goodness-of-fit**.
- The question ***"should we trust more the tests or the p-value?"*** aroused during estimation of hybrid Gibbs process model explaining location of firms, when goodness-of-fit check (QQplot and G residual curve) confirmed the correctness of new applied method, but the key parameters (interaction terms) were not statistically significant:

1. Aim of the study: How it started

Coefficient	Value and significance
Intercept	-5.18*
Roads	-0.03
Centre	-0.31**
Interaction_1	1.57
Interaction_2	0.05

Note: significance for 10% is denoted with dot, for 5% - with *, for 1% - with **

1. Aim of the study: How it started

Coefficient	Value and significance
Intercept	-5.18*
Roads	-0.03
Centre	-0.31**
Interaction_1	1.57
Interaction_2	0.05

Note: significance for 10% is denoted with dot, for 5% - with *, for 1% - with **

1. Aim of the study: How it started

Coefficient	Value and significance
Intercept	-5.18*
Roads	-0.03
Centre	-0.31**
Interaction_1	1.57
Interaction_2	0.05

Note: significance for 10% is denoted with dot, for 5% - with *, for 1% - with **

Interaction terms **are not significant**, although goodness-of-fit analysis **shows the correctness of interaction form**.

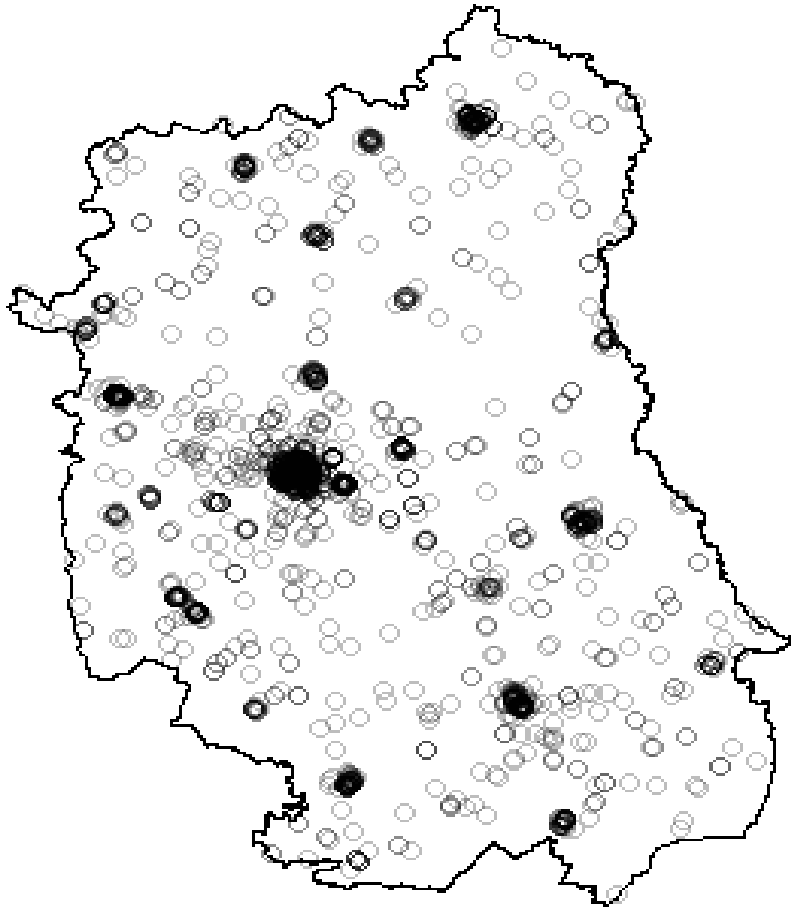
1. Aim of the study: Summary

- Research aims to check **how statistical significance of estimated parameters is connected with model's goodness-of-fit**.
- The question ***"should we trust more the tests or the p-value?"*** aroused during estimation of hybrid Gibbs process model explaining location of firms, when goodness-of-fit check (QQplot and G residual curve) confirmed the correctness of new applied method, but the key parameters (interaction terms) were not statistically significant.
- Thus, we needed **a statistical evidence that a new approach** (Baddeley-Geyer process model) **is better than an old one** (single Geyer process model).

2. Theoretical issues

Point pattern

- In a very simple words, point pattern is a set of points in two-dimensional space. Those points can overlap (have the same location). Each point pattern is bounded with observation region (window), of regular/irregular shape
- Most common point patterns are random, clustered and regular
- Each point inside pattern can possess some feature – categorical or numerical



Source: own illustration of real data point pattern – sample of firms from Lubelskie voivodeship

2. Theoretical issues

- Gibbs models were primarily defined to examine interaction in biology (between different species)
- However, as firms compete also for territory, resources and clients, those processes can be applied in firm location studies
- Several Gibbs processes cannot be applied for business related studies due to:

1. Their application field is different from economics
2. They were created for specifically regular patterns
3. They describe hierarchy between points, which is not applicable in a case of examining an entry of new firm, which is not subsidiary for any of already existing on the market
4. Their estimation requires choice of irregular parameter

+ firms rather interact with more than 1 neighbour => This implies choice of (Baddeley-)Geyer saturation model

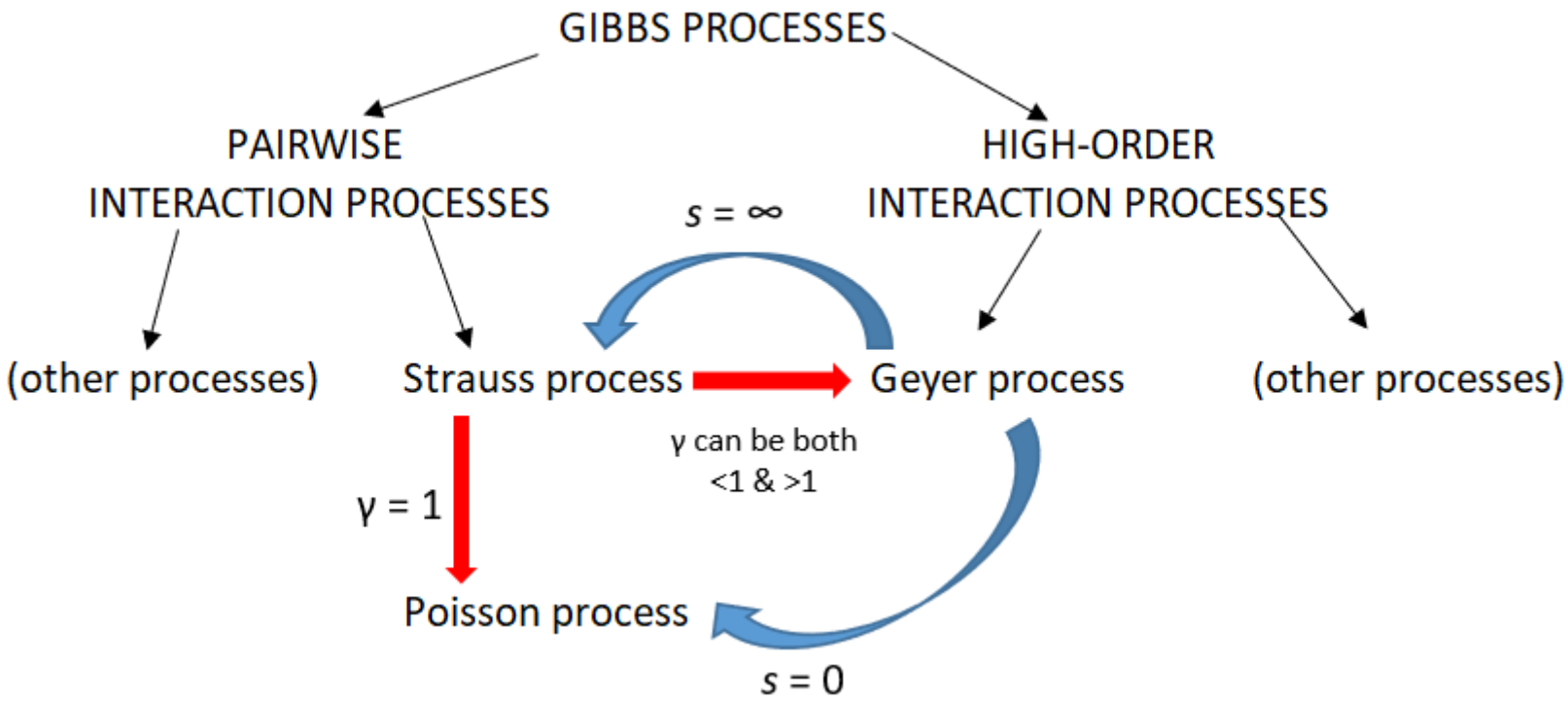
2. Theoretical issues

- For this model, two parameters are important:

1. Interaction radius – radius, within which interaction between points exists
2. Saturation parameter – "*overall contribution from each point to the total interaction is trimmed to never exceed a maximum value*" (Baddeley et al. 2015, Ch.13.7.2)

In our case, we assume more than one radius within which interaction exists

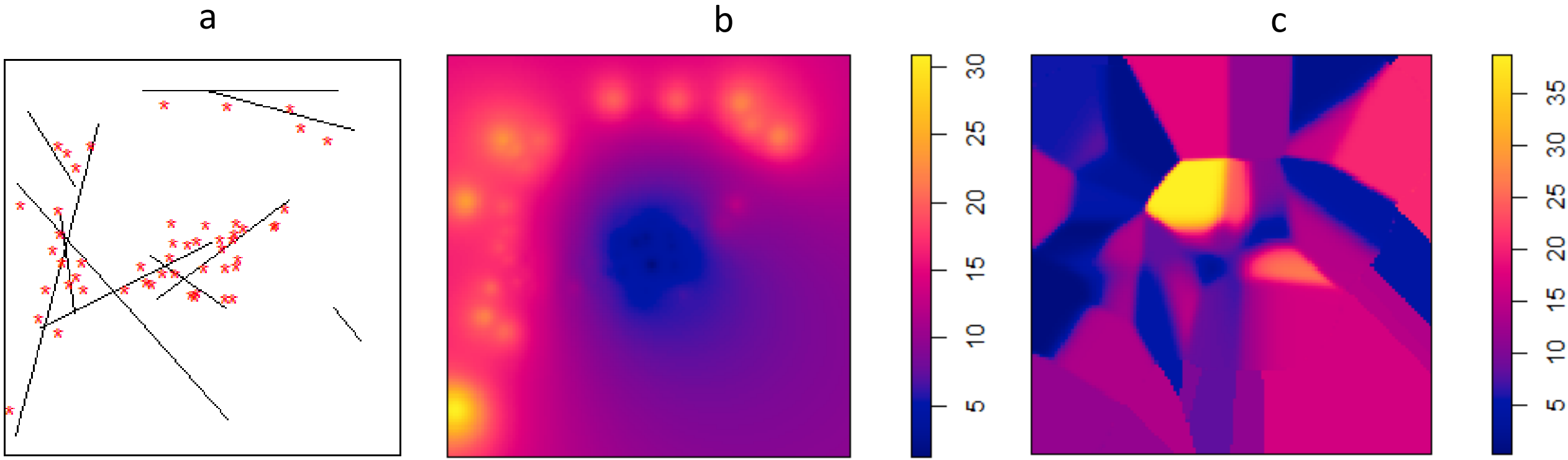
2. Theoretical issues



$$\lambda(u, \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x} \setminus \{u\})}$$ (Baddeley et al., 2013, p.6)

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \prod_{i=1}^{n(\mathbf{x})} \gamma^{\min(s, t(x_i, \mathbf{x} \setminus x_i))}$$ (Baddeley et al., 2013, p. 8)

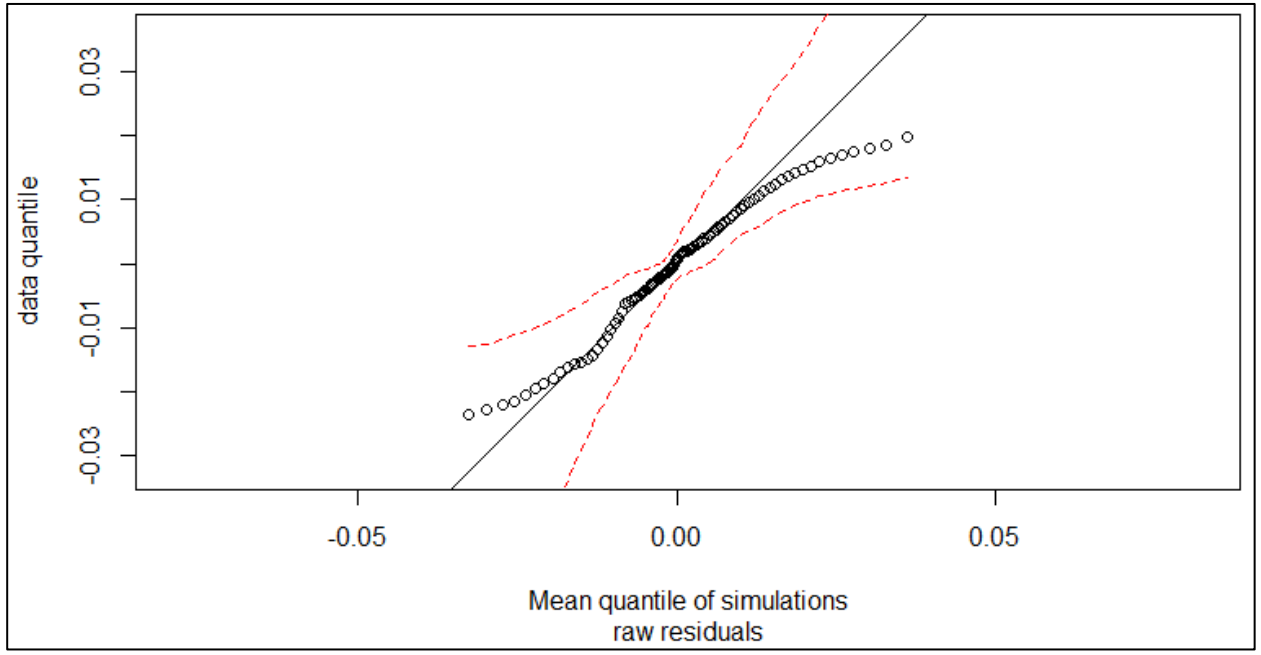
3. Data and covariates

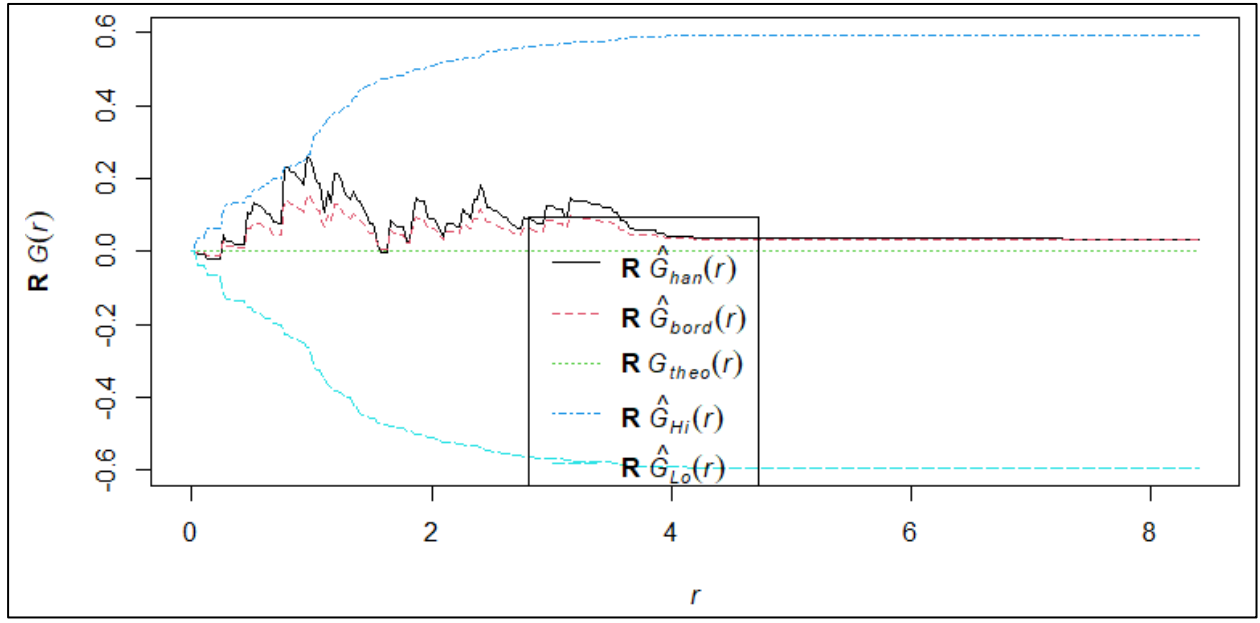


Study pattern of points with line pattern of roads (a), distance from each point to the centre (b), distance from each point to line pattern of roads (c)

4. Estimation results

Is chosen interpoint interaction correct? -> Yes, no outliers for qqplot and curve within bound for Gres





4. Estimation results

- Is chosen interpoint interaction correct? -> Yes, no outliers for qqplot and curve within bound for Gres
- The **interaction form of the model is correct** what suggest that it is indeed the Baddeley-Geyer process, but on the other hand, **interaction terms are not significant**.
- This raises two questions – (a) **should we trust more the tests or the p-value?** and (b) **is there a connection between statistical significance and goodness-of-fit?**

5. What can single hypothesis tell us?

- Analysis starts with consideration of a single hypothesis about the insignificance of a parameter. In such situation, we have the following probabilities of type I and type II error:

		Test concludes	
	Hypothesis	H0: $\gamma_i = 0$	H1: $\gamma_i \neq 0$
Truth	H0: $\gamma_i = 0$	$1 - \alpha$	α
	H1: $\gamma_i \neq 0$	β	$1 - \beta$

5. What can single hypothesis tell us?

- Analysis starts with consideration of a single hypothesis about the insignificance of a parameter. In such situation, we have the following probabilities of type I and type II error:

		Test concludes	
	Hypothesis	H0: $\gamma_i = 0$	H1: $\gamma_i \neq 0$
Truth	H0: $\gamma_i = 0$	$1 - \alpha$	α
	H1: $\gamma_i \neq 0$	β	$1 - \beta$

- Focusing on the situation, where test concludes rejection of H0 in favour of H1, one can analyse the relationship between the rejection of the null hypothesis and the values of alfa and beta.

5. What can single hypothesis tell us?

		Test concludes	
	Hypothesis	H0: $\gamma_i = 0$	H1: $\gamma_i \neq 0$
Truth	H0: $\gamma_i = 0$	$1 - \alpha$	α
	H1: $\gamma_i \neq 0$	β	$1 - \beta$

- Focusing on the situation, where test concludes rejection of H0 in favour of H1, one can analyse the relationship between the rejection of the null hypothesis and the values of alfa and beta.
- If one assumes that one may equally probably reject or not reject the null, then, for two independently performed tests (for two interaction parameters), probability, that both tests will reject the null hypothesis is:

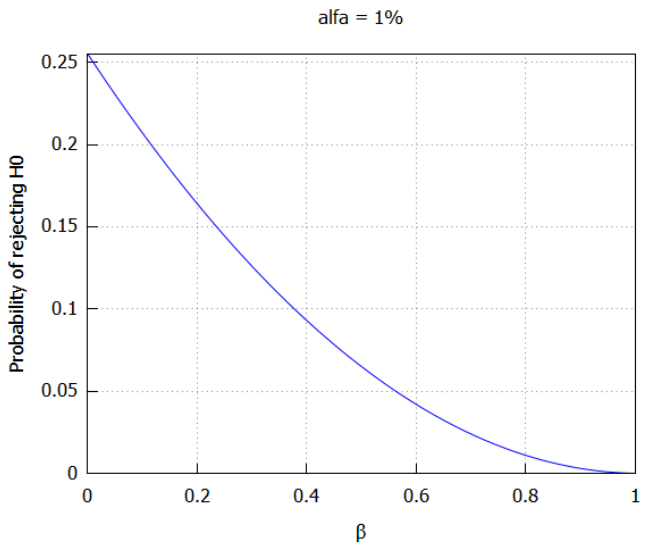
$$\mathbb{P} = \frac{1}{2} * (1 - \beta + \alpha) * \frac{1}{2} * (1 - \beta + \alpha) = \frac{1}{4} * (1 - 2\beta + 2\alpha - 2\alpha\beta + \alpha^2 + \beta^2)$$

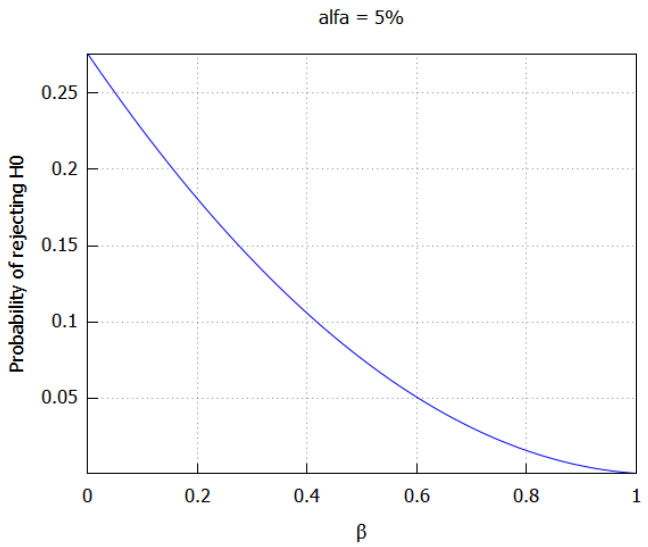
5. What can single hypothesis tell us?

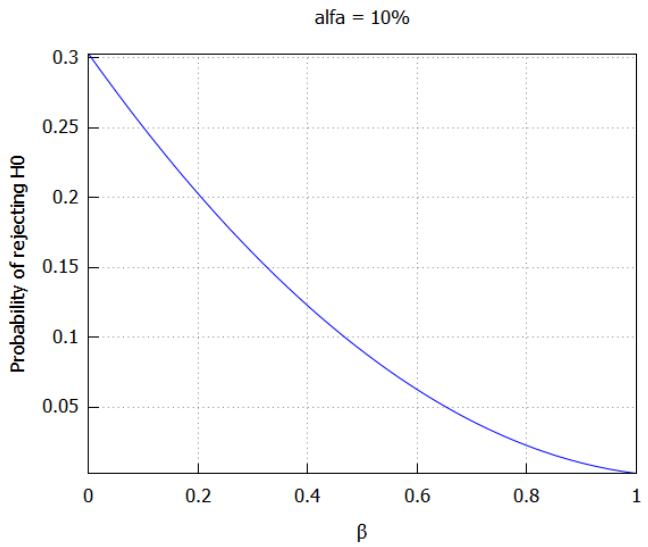
- For fixed alfa (1%-5%-10%), probability, that both tests will reject the null hypothesis can be represented as follows:

5. What can single hypothesis tell us?

- For fixed alfa (1%-5%-10%), probability, that both tests will reject the null hypothesis can be represented as follows:







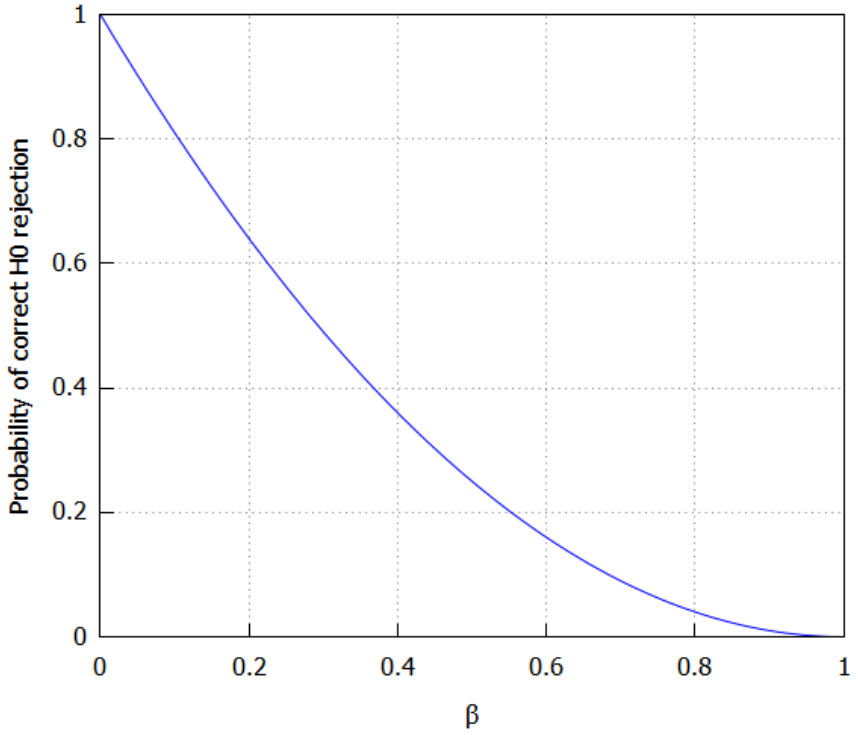
- The higher the beta, the less probable that the test rejects the null about insignificance (without considering whether such a rejection is true or false)

5. What can single hypothesis tell us?

- In case where an incorrect null hypothesis is **correctly** rejected, such probability equals to $(1-\beta) * (1-\beta) = 1 - 2\beta + \beta^2$, which gives the following graph:

5. What can single hypothesis tell us?

- In case where an incorrect null hypothesis is **correctly** rejected, such probability equals to $(1-\beta) * (1-\beta) = 1 - 2\beta + \beta^2$, which gives the following graph:



6. Simulation results

- In order to test how the significance of interaction terms corresponds to the results of goodness-of-fit tests, **100 simulations** were made
- For each of them, it was analysed in **how many cases** and **which** interaction/-s **was/were found to be significant**
- It was checked also, **in how many cases QQ-plot and G-residual curve suggested correctness of spatial interaction**.

6. Simulation results

- In order to test how the significance of interaction terms corresponds to the results of goodness-of-fit tests, **100 simulations** were made
- For each of them, it was analysed in **how many cases** and **which** interaction/-s **was/were found to be significant**
- It was checked also, **whether QQ-plot and G-residual curve suggested correctness of spatial interaction**.
- *Additionally*, performance of old and new approached were compared with values of AIC and analysis of coefficients' distributions.*

6. Simulation results

The probability, that both tests will reject the null hypothesis:

$$\mathbb{P} = \frac{1}{2} * (1 - \beta + \alpha) * \frac{1}{2} * (1 - \beta + \alpha) = \frac{1}{4} * (1 - 2\beta + 2\alpha - 2\alpha\beta + \alpha^2 + \beta^2)$$

Case	Rejection of null for γ_1	Rejection of null for γ_2	Rejection for both γ_1 and γ_2	Non-rejection for both γ_1 and γ_2
Percent of cases	48%	7%	34%	11%

Note: results are given for 91 models; among 100 patterns, for 2 of them algorithm couldn't find the optimal saturation parameter and radius which were about to be put into a model, and for 7 more models estimates were not found because of lack of pseudo-likelihood maximization convergence

Source: own calculations in spatstat::

6. Simulation results

The probability, that both tests will reject the null hypothesis:

$$\mathbb{P} = \frac{1}{2} * (1 - \beta + \alpha) * \frac{1}{2} * (1 - \beta + \alpha) = \frac{1}{4} * (1 - 2\beta + 2\alpha - 2\alpha\beta + \alpha^2 + \beta^2)$$

Case / Graph	Correct QQ-plot	Correct G-residual plot	Both plots are correct
All models	69%	76%	59%
Models with two significant interaction parameters	77%	90%	71%

Note: results are given for 91 models; among 100 patterns, for 2 of them algorithm failed to find an optimal saturation parameter and radius which were about to put into a model, and for 7 more models estimates were not found because of lack of pseudo-likelihood maximization convergence

Source: own calculations in spatstat::

7. Implications and conclusions

- Proposed hybrid Gibbs process model (used to analyse location of firms) **has statistically proven to perform better than older known approach**.
- Both interactions were significant in 1/3 of cases, but **only in 11% of cases they were not statistically different from zero**.
- Additionally, **in 71% of cases, the two-radii model** (hybrid Gibbs process) **has the correct interaction form**. The incorrectness of interaction for other models can be explained by the fact that the true underlying process is even more complicated and cannot be easily determined.
- Taking into account the fact, that parameters for the hybrid process are not chosen totally by an algorithm, such results can be considered very good ones. **They create a possibility for further analysis of business location with point process methods**.

Thank you for your attention!

Do not hesitate to ask questions ☺

 k_zabarina

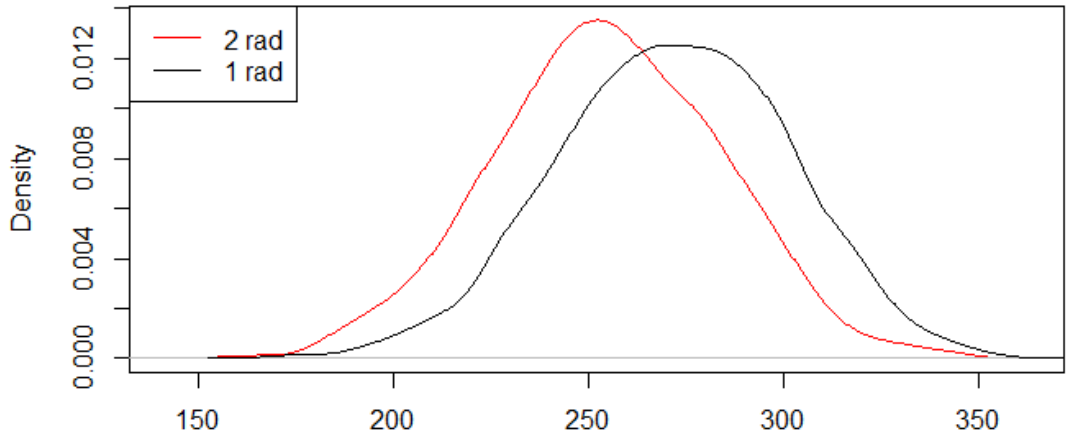
 SpatialWarsaw

8. Additional slides*

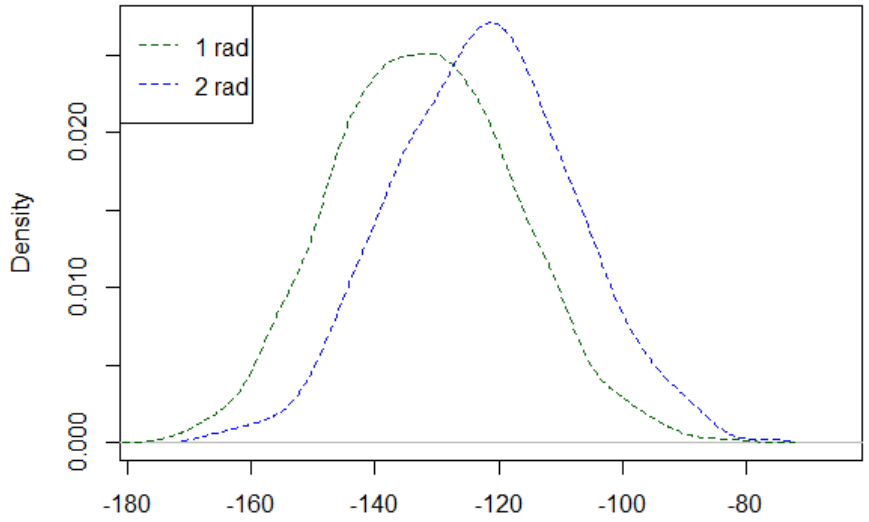
A1. AIC and pseudo loglik, 1 vs 2 radii

	AIC		Pseudo loglik	
	2 radii	1 radius	2 radii	1 radius
Min.	173.0	191.4	-159.88	-173.14
1st Qu.	234.9	257.4	-134.04	-145.75
Median	257.6	280.0	-123.81	-135.98
Mean	256.1	278.6	-123.07	-135.30
3rd Qu.	278.1	299.5	-112.44	-124.72
Max.	329.8	354.3	-81.51	-91.72

Comparison of AIC (a) and pseudo-likelihood (b) densities for models with one and two radii



(a)



(b)